An algebraic formulation of the graph reconstruction conjecture

Igor C. Oliveira *
Columbia University
oliveira@cs.columbia.edu

Bhalchandra D. Thatte †
Universidade de São Paulo
bdthatte@ime.usp.br

January 18, 2013

Abstract

The graph reconstruction conjecture asserts that a finite simple graph on at least 3 vertices can be reconstructed up to isomorphism from its deck - the collection of its vertex-deleted subgraphs. Kocay's Lemma is an important tool in graph reconstruction. Roughly speaking, given the deck of a graph G and any finite sequence of graphs, it gives a linear constraint that every reconstruction of G must satisfy.

Let $\psi(n)$ be the number of distinct (mutually non-isomorphic) graphs on n vertices, and let $d(n)$ be the number of distinct decks that can be constructed from these graphs. Then the difference $\psi(n) - d(n)$ is a measure of how many graphs cannot be reconstructed from their decks. In particular, the graph reconstruction conjecture holds for n-vertex graphs if and only if $\psi(n) = d(n)$. We give a framework based on Kocay's lemma to study this discrepancy. We prove that if M is a matrix of covering numbers of graphs by sequences of graphs then $d(n) \geq \mathsf{rank}_{\mathbb{R}}(M)$. In particular, all n-vertex graphs are reconstructible if one such matrix has rank $\psi(n)$. To complement this result, we prove that it is possible to choose a family of sequences of graphs such that the corresponding matrix M of covering numbers satisfies $d(n) = \mathsf{rank}_{\mathbb{R}}(M)$.

1 Introduction

The graph reconstruction conjecture was proposed by Ulam [11] and Kelly [5]. Informally, it states that if two finite, undirected, simple graphs on at least 3 vertices have the same collection (*multi-set* or *deck*) of *unlabelled* vertex-deleted subgraphs, then the graphs are isomorphic; in other words, any such graph can be *reconstructed* up to isomorphism from the collection of its unlabelled vertex-deleted subgraphs.

The conjecture has been verified by McKay [8] for all undirected, finite, simple graphs on eleven or fewer vertices. In addition, it has been proven for many particular classes of undirected, finite, simple graphs, such as regular graphs, disconnected graphs and trees (Kelly [6]). In fact, Bollobás [2] showed that it holds for almost all finite, simple, undirected graphs. On the other hand, a similar conjecture does not hold for directed graphs: Stockmeyer [9, 10] constructed a number of

*Supported in part by NSF grants CCF-0915929 and CCF-1115703.
†Supported by CNPq grant 151782/2010-5 and by MaCLinC Project at Universidade de São Paulo.

infinite families of non-reconstructible directed graphs. For a more comprehensive introduction to the problem, we refer to a survey by Bondy [3]. For the standard graph theoretic terminology not defined here, we refer to West [12].

Kelly's Lemma [6] is one of the most useful results in graph reconstruction. Let $s(F, G)$ denote the number of subgraphs of G isomorphic to F. Kelly's lemma states that for $v(F) < v(G)$, the parameter $s(F, G)$ is reconstructible, in the sense that if G and G' have the same deck then $s(F, G') = s(F, G)$. Several propositions in graph reconstruction rely on this useful lemma.

Kocay's Lemma [7] allows us, to some extent, to overcome the restriction $v(F) < v(G)$ in Kelly's lemma. It provides a linear constraint on $s(F, G)$ that must be satisfied by every reconstruction of G. Informally, it says that, if $\mathcal{F} = (F_1, \ldots, F_m)$ is a sequence of graphs, each of which has at most $v(G) - 1$ vertices, then there are constants $c(\mathcal{F}, H)$ such that the value of the sum $\sum_H c(\mathcal{F}, H) \cdot s(H, G)$ is reconstructible, where the sum is taken over all unlabelled n-vertex graphs H. Roughly speaking, the constant $c(\mathcal{F}, H)$ counts the number of ways to *cover* the graph H by graphs in the sequence $\mathcal{F}$.

Kocay's Lemma has been used to show several interesting results in graph reconstruction. For instance, by carefully selecting the sequence $\mathcal{F}$, it is possible to give a simple proof that disconnected graphs are reconstructible. In addition, it can be used to show that the number of perfect matchings, the number of spanning trees, the characteristic polynomial, the chromatic polynomial, and many other parameters of interest are reconstructible; see Bondy [3].

It is natural to wonder whether even more restrictions may be imposed on the reconstructions of G by applications of Kocay's Lemma. Recall that it is possible to use different sequences of graphs in each such invocation of Kocay's lemma, and as explained before, for each such sequence we get a linear constraint that the reconstructions of G must satisfy. By analysing such equations one would expect to obtain a wealth of information about the structure of any reconstruction of G (perhaps enough equations may even allow us to conclude that G is reconstructible). In this paper we investigate how much information one can obtain by setting up such equations.

We prove that the equations obtained by applying Kocay's Lemma to the deck of a graph G using distinct sequences of graphs provide important information not only about the reconstructions of G, but also about the total number of non-reconstructible graphs on n vertices. More formally, let $d(n)$ be the number of distinct decks obtained from n-vertex graphs. We show that if M is the matrix of coefficients corresponding to these equations, then $d(n) \geq \mathsf{rank}_{\mathbb{R}}(M)$, i.e., the rank of this matrix provides a lower bound on the number of distinct decks. In particular, the existence of a full-rank matrix of coefficients would imply that all graphs on n vertices are reconstructible. In addition, we give a proof that there exist $d(n)$ sequences of graphs $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_{d(n)}$, with corresponding matrix M of covering numbers, such that $\mathsf{rank}_{\mathbb{R}}(M) = d(n)$. In other words, if the graph reconstruction conjecture holds for graphs with n vertices, then there is a corresponding full-rank matrix certifying this fact.

We state our results in more generality for graphs, hypergraphs, directed graphs, and also for classes of graphs for which similar equations can be constructed; for example, analogous results hold for planar graphs, disconnected graphs and trees.

2 Preliminaries

In this paper, we consider general finite graphs - undirected graphs, directed graphs, hypergraphs, graphs with or without multiple edges, and with or without loops. We take the vertex set

of a graph to be a finite subset of $\mathbb{N}$. We write $V^{(k)}$ for the family of k-element subsets of a set V.

Definition 2.1 (Graphs)**.** A *hypergraph* G is a triple (V, E, ϕ), where V is its *vertex set* (also called *ground set*, and written as $V(G)$) and E is its set of *hyperedges* (written as $E(G)$), and a map $\phi : E \to 2^V \setminus \emptyset$[1]. An *undirected graph* G is a hypergraph with the restriction that $\phi : E \to V^{(1)} \cup V^{(2)}$; in this case we call a hyperedge e an *edge* (if $|\phi(e)| = 2$) or a *loop* (if $|\phi(e)| = 1$). A *directed graph* G is a triple (V, E, ψ), where V is its vertex set and E is the set of its *arcs*, and a map $\psi : E \to V \times V$. The first element of $\psi(e)$ is called the *tail* of the arc e, and the second element of $\psi(e)$ is called the *head* of e. We denote the set of all finite graphs (including hypergraphs, undirected graphs and directed graphs) by $\mathcal{G}^*$.

Remark 2.2. Although our results and proofs are stated in full generality, it may be helpful in a first reading to consider only finite, simple, undirected graphs.

Definition 2.3 (Graph isomorphism)**.** Let G and H be two graphs. We say that G and H are *isomorphic* (written as $G \cong H$) if there are one-one maps $f : V(G) \to V(H)$ and $g : E(G) \to E(H)$ such that an edge e and a vertex v are incident in G if and only the edge $g(e)$ and the vertex $f(v)$ are incident in H. Additionally, in the case of directed graphs, a vertex v is the head (or the tail) of an arc e if and only if $f(v)$ is the head (or, respectively, the tail) of $g(e)$. The isomorphism class of a graph G, denoted by $G/\!\cong$, is the set of graphs isomorphic to G.

Definition 2.4. A *class of graphs* is a set of graphs that is closed under isomorphism. A class of graphs is said to be *finite* if contains finitely many isomorphism classes.

Definition 2.5 (Reconstruction)**.** Let G be graph and let v be a vertex of G. The induced subgraph of G obtained by deleting v and all edges incident with v is called a *vertex-deleted subgraph* of G, and is written as $G - v$. We say that H is a *reconstruction* of G (written as $H \sim G$) if there is a one-one map $f : V(G) \to V(H)$ such that for all $v \in V(G)$, the graphs $G - v$ and $H - f(v)$ are isomorphic. The relation $\sim$ is an equivalence relation. We say that a graph G is *reconstructible* if every reconstruction of G is isomorphic to G (i.e., if $H \sim G$ implies $H \cong G$). A parameter $t(G)$ is said to be *reconstructible* if $t(H) = t(G)$ for all reconstructions H of G. Let $\mathcal{C}$ be a class of graphs. We say that $\mathcal{C}$ is *recognisable* if, for any $G \in \mathcal{C}$, every reconstruction of G is in $\mathcal{C}$. Furthermore, we say that $\mathcal{C}$ is *reconstructible* if every graph $G \in \mathcal{C}$ is reconstructible.

Example 2.6. Let $G(V, E, \phi)$ be a hypergraph. The number of edges incident with all vertices (i.e., edges $e \in E$ such that $\phi(e) = V$, which we call *big edges*), is not a reconstructible parameter. For example, if G^k is a graph obtained from G by adding k new edges $e_1, e_2, \ldots, e_k$ and making them incident with all vertices in V, then G^k is a reconstruction of G. In this sense, no hypergraphs are reconstructible, and each hypergraph has infinitely many mutually non-isomorphic reconstructions. If G is a graph in class $\mathcal{C}$, then $\mathcal{C}$ is not recognisable if for some k, the graph G^k is not in $\mathcal{C}$; and $\mathcal{C}$ is not finite if graphs G^k are all in $\mathcal{C}$. On the other hand, the number of *small edges*, i.e., edges $e \in E$ such that $\phi(e) \neq V$, is a reconstructible parameter.

In view of the above example, we will always use $\mathcal{G}^*$ for the set of all graphs, $\mathcal{G}$ for the set of all graphs without big edges, and $\mathcal{G}_n$ for the set of n-vertex graphs without big edges. A class $\mathcal{C}_n$ will always be a subset of $\mathcal{G}_n$. We will use the following slightly restrictive definitions for some other reconstruction terms.

[1] Observe that we are defining graphs using triples because multiple edges are allowed.

Definition 2.7. A graph G in $\mathcal{G}$ is reconstructible if it is *reconstructible modulo big edges*, i.e., if G' is a reconstruction of G and $G' \in \mathcal{G}$, then G' is isomorphic to G. A subclass $\mathcal{C}$ of $\mathcal{G}$ is recognisable if for each graph G in $\mathcal{C}$, each reconstruction of G in $\mathcal{G}$ is also in $\mathcal{C}$. A subclass $\mathcal{C}$ of $\mathcal{G}$ is reconstructible if each graph in $\mathcal{C}$ is reconstructible (modulo big edges).

Example 2.8. Disconnected undirected graphs on 3 or more vertices are recognisable and reconstructible. However, there are classes of graphs that are recognisable, but not known to be reconstructible. An important example is the class of planar graphs (Bilinski et al. [1]).

Since $\cong$ and $\sim$ are equivalence relations, the *quotient* notation may be conveniently used to define various equivalence classes of graphs. We write the set of all isomorphism classes of graphs as $\mathcal{G}^*/\cong$; analogously we use $\mathcal{G}_n/\cong$, $\mathcal{C}/\cong$, $\mathcal{C}_n/\cong$, and so on. We define an *unlabelled graph* to be an isomorphism class of graphs. But sometimes we abuse the notation slightly, e.g., if a quantity is invariant over an isomorphism class H, then in the same context we may also use H to mean a representative graph in the class. Similarly, we denote various reconstruction classes by $\mathcal{G}/\sim$, $\mathcal{G}_n/\sim$, $\mathcal{C}/\sim$, $\mathcal{C}_n/\sim$, and so on. Note that equivalence classes of any class of graphs under $\sim$ are refined by $\cong$; in particular, $|\mathcal{C}_n/\sim| \le |\mathcal{C}_n/\cong|$, and equality holds if and only if the class $\mathcal{C}_n$ is reconstructible. We will refer to reconstruction classes of $\mathcal{C}_n$ (i.e., members of $\mathcal{C}_n/\sim$) by $R_1, R_2, \dots$, and isomorphism classes of R_i (i.e., members of $R_i/\cong$) by $R_{i,1}, R_{i,2}, \dots$.

Given graphs G and H, the number of subgraphs of G isomorphic to H is denoted by $s(H, G)$. The following two subgraph counting lemmas are important results about the reconstructibility of the parameter $s(H, G)$.

Lemma 2.9 (Kelly's lemma, [6])**.** *Let H be a reconstruction of G. If F is any graph such that $v(F) < v(G)$, then $s(F, G) = s(F, H)$.*

Definition 2.10. Let G be a graph and let $\mathcal{F} \coloneqq (F_1, F_2, \dots, F_m)$ be a sequence of graphs. A *cover* of G by $\mathcal{F}$ is a sequence $(G_1, G_2, \dots, G_m)$ of subgraphs of G such that $G_i \cong F_i$, $1 \le i \le m$, and $\bigcup G_i = G$. The number of covers of G by $\mathcal{F}$ is denoted by $c(\mathcal{F}, G)$.

Lemma 2.11 (Kocay's lemma, *[7]*)**.** *Let G be a graph on n vertices. For any sequence of graphs $\mathcal{F} \coloneqq (F_1, F_2, \dots, F_m)$, where $v(F_i) < n$, $1 \le i \le m$, the parameter*

$$\sum_H c(\mathcal{F}, H) s(H, G)$$

is reconstructible, where the sum is over all unlabelled n-vertex graphs H.

Proof. We count in two ways the number of sequences $(G_1, \dots, G_m)$ of subgraphs of G such that $G_i \cong F_i$, $1 \le i \le m$. We have

$$\prod_{i=1}^{m} s(F_i, G) = \sum_X c(\mathcal{F}, X) s(X, G), \tag{1}$$

where the sum extends over all unlabelled graphs X on at most n vertices. Since $v(F_i) < n$, it follows by Kelly's Lemma that the left-hand side of this equation is reconstructible. On the other hand, the terms $c(\mathcal{F}, X)s(X, G)$ are also reconstructible whenever $v(X) < n$. The result follows after rearranging Equation 1. $\square$

To state our results in full generality, we make the following definition.

Definition 2.12. Let $\mathcal{C}_n$ be a class of graphs on n vertices. We say that $\mathcal{C}_n$ *satisfies Kocay's lemma* if, for every graph $G \in \mathcal{C}_n$ and every sequence of graphs $\mathcal{F} = (F_1, F_2, \ldots, F_m)$, where $v(F_i) < n$, $1 \leq i \leq m$, the sum

$$\sum_{H \in \mathcal{C}_n/\cong} c(\mathcal{F}, H)s(H, G)$$

is reconstructible.

The following proposition gives a simple condition that is sufficient for a class of graphs $\mathcal{C}_n$ to satisfy Kocay's lemma.

Proposition 2.13. *Let $\mathcal{C}_n$ be a class of graphs on n vertices. Suppose that $s(H, G)$ is reconstructible for every $G \in \mathcal{C}_n$ and for every n-vertex graph $H \notin \mathcal{C}_n$. Then the class $\mathcal{C}_n$ satisfies Kocay's lemma.*

Proof. Let $G \in \mathcal{C}_n$. Let $\mathcal{F} := (F_1, F_2, \ldots, F_m)$ be any sequence of graphs such that $v(F_i) < n$, $1 \leq i \leq m$. We write the R.H.S. of Equation 1 as

$$\sum_{H \in \mathcal{C}_n/\cong} c(\mathcal{F}, H)s(H, G) + \sum_{H \notin \mathcal{C}_n/\cong} c(\mathcal{F}, H)s(H, G),$$

where the second summation is reconstructible. Now we rearrange the terms in Equation 1 to obtain $\sum_{H \in \mathcal{C}_n/\cong} c(\mathcal{F}, H)s(H, G)$. □

The class of connected simple graphs satisfies Kocay's lemma since if G is any connected graph and H is any disconnected graph, then $s(H, G)$ is reconstructible (see Bondy [3]). Other classes of graphs that satisfy Kocay's lemma include planar graphs, trees and of course the class of all graphs. Our theorems apply to finite and recognisable classes of graphs satisfying Kocay's Lemma. All the above classes of graphs are recognisable as well.

Let $\mathcal{C}_n \subseteq \mathcal{G}_n$ be a finite, recognisable class of n-vertex graphs satisfying Kocay's Lemma. In the rest of this paper, we study equations obtained by applying Kocay's Lemma to $\mathcal{C}_n$. It is useful to view this lemma as follows. Let $\mathcal{F} := (F_1, \ldots, F_m)$, be a sequence of graphs where $v(F_i) < n$ for each $1 \leq i \leq m$. Let $G, G' \in R \in \mathcal{C}_n/\sim$, i.e., G' is a reconstruction of G, and since $\mathcal{C}_n$ is recognisable, G' is in $\mathcal{C}_n$. Then we have

$$\sum_{H \in \mathcal{C}_n/\cong} c(\mathcal{F}, H)s(H, G') = k_{\mathcal{F},R},$$

where $k_{\mathcal{F},R}$ is a constant that depends only on the sequence $\mathcal{F}$ and the reconstruction class R, i.e., it is a reconstructible parameter. In this expression, $c(\mathcal{F}, H)$ is constant (i.e., it is independent of the reconstruction class) and $s(H, G')$ depends on the isomorphism class of a particular reconstruction G' of G under consideration. Therefore, each application of Kocay's Lemma provides a linear constraint on $s(H, G')$ that all reconstructions G' of G must satisfy.

This paper is devoted to a study of systems of such linear constraints obtained by applications of Kocay's lemma. In particular, we study the rank of a matrix of covering numbers that we define next.

Definition 2.14. Let $\mathcal{C}_n$ be a finite class of graphs on n vertices. Let $\mathfrak{F} = (\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_l)$ be a family of sequences of graphs on at most $n-1$ vertices. We let $M_{\mathfrak{F}, \mathcal{C}_n/\cong} \in \mathbb{R}^{|\mathfrak{F}| \times |\mathcal{C}_n/\cong|}$ to be a matrix whose rows are indexed by the sequences $\mathcal{F}_i, i = 1, 2, \ldots, l$ and whose columns indexed by the distinct isomorphism classes of graphs in $\mathcal{C}_n$. The entries of $M_{\mathfrak{F}, \mathcal{C}_n/\cong}$ are the covering numbers defined by $c(\mathcal{F}, H)$, where $\mathcal{F} \in \mathfrak{F}$ and $H \in \mathcal{C}_n/\cong$.

3 On the rank of a matrix obtained from Kocay's Lemma

3.1 Large rank implies few non-reconstructible graphs

As observed earlier, for any finite class $\mathcal{C}_n$ of graphs, $|\mathcal{C}_n/\sim| \leq |\mathcal{C}_n/\cong|$, and the bigger the number of distinct reconstruction classes, the smaller is the number of non-reconstructible graphs. The main result Theorem 3.2 of this section states that for any finite, recognisable class of graphs satisfying Kocay's lemma, the number of distinct reconstruction classes is bounded below by the rank of the matrix of covering numbers for any system of sequences of graphs.

Let $\mathcal{C}_n$ be a finite, recognisable class of n-vertex graphs satisfying Kocay's Lemma. Let $\mathfrak{F}$ be a finite family of sequences of graphs on at most $n-1$ vertices. Let $M_{\mathfrak{F}, \mathcal{C}_n/\cong}$ be the corresponding matrix of covering numbers $c(\mathcal{F}, H)$, where $\mathcal{F} \in \mathfrak{F}$ and $H \in \mathcal{C}_n/\cong$ (see Definition 2.14). Let $W = \{x \in \mathbb{R}^{|\mathcal{C}_n/\cong|} \mid M_{\mathfrak{F}, \mathcal{C}_n/\cong} \cdot x \equiv 0\}$ be a subspace of the vector space $\mathbb{R}^{|\mathcal{C}_n/\cong|}$ over $\mathbb{R}$. We associate with $\mathcal{C}_n$ the constant $\alpha(\mathcal{C}_n) := |\mathcal{C}_n/\cong| - |\mathcal{C}_n/\sim|$.

Lemma 3.1. $\mathsf{dim}(W) \geq \alpha(\mathcal{C}_n)$.

Proof. If $\alpha(\mathcal{C}_n) = 0$, the result is trivial. Otherwise, let $R_1, \ldots, R_s \in \mathcal{C}_n/\sim$ be the non-reconstructible reconstruction classes in $\mathcal{C}_n$, i.e., $r_i := |R_i/\cong| > 1$ for all $i \in \{1, 2, \ldots, s\}$. Let $R_{i,j}, j \in \{1, 2, \ldots, r_i\}$ be the isomorphism classes in $R_i, i \in \{1, 2, \ldots, s\}$. Let $G_{i,j}$ be representative graphs from $R_{i,j}$.

For each $G_{i,j}$, we define a vector $w^{i,j} \in \mathbb{R}^{|\mathcal{C}_n/\cong|}$, with its entries, which are indexed by unlabelled graphs $H \in \mathcal{C}_n/\cong$, defined as follows:

$$w^{i,j}(H) := s(H, G_{i,j}) - s(H, G_{i,1}), \text{ where } H \in \mathcal{C}_n/\cong .$$

Observe that to prove the lemma it is enough to show that the vectors $w^{i,j}$ satisfy the following properties:

(*i*) for all $i \in \{1, 2, \ldots, s\}$, for all $j \in \{1, 2, \ldots, r_i\}$, $w^{i,j} \in W$; and

(*ii*) the vectors in the set $U := \{w^{i,j} \mid 1 \leq i \leq s, 2 \leq j \leq r_i\}$ are non-zero and linearly independent, where $|U| = \alpha(\mathcal{C}_n)$.

Proof of (*i*): Graphs $G_{i,j}$ and $G_{i,1}$ are reconstructions of each other, and $\mathcal{C}_n$ satisfies Kocay's Lemma. Therefore, for every row $M_{\mathcal{F}}$ of $M_{\mathfrak{F}, \mathcal{C}_n/\cong}$, we have,

$$\begin{aligned} \sum_{H \in \mathcal{C}_n/\cong} c(\mathcal{F}, H) s(H, G_{i,j}) &= \sum_{H \in \mathcal{C}_n/\cong} c(\mathcal{F}, H) s(H, G_{i,1}) \\ \therefore \quad M_{\mathcal{F}} \cdot w^{i,j} &= \sum_{H \in \mathcal{C}_n/\cong} c(\mathcal{F}, H)(s(H, G_{i,j}) - s(H, G_{i,1})) = 0. \end{aligned}$$

Therefore, $M_{\mathfrak{F}, \mathcal{C}_n/\cong} \cdot w^{i,j} = 0$.

Proof of (ii): Let the vectors in U be ordered $u^1, u^2, \ldots, u^{\alpha(\mathcal{C}_n)}$ so that the corresponding graphs are ordered by non-decreasing numbers of small edges. We prove that u^1 is non-zero, and for each $k \in \{2, \ldots, \alpha(\mathcal{C}_n)\}$, the vector u^k is non-zero and is linearly independent of $u^1, u^2, \ldots, u^{k-1}$, which would imply that the vectors in U are linearly independent.

Let $u^\ell = w^{i,j}$ for some $i \in \{1, 2, \ldots, s\}$ and $j \in \{2, \ldots, r_i\}$. First recall that $\mathcal{C}_n$ is recognisable, $R_i \in \mathcal{C}_n/\!\sim$, and $G_{i,j} \in R_i/\!\cong$; therefore, $G_{i,j} \in \mathcal{C}_n/\!\cong$. In addition, $G_{i,j} \ncong G_{i,1}$ since $j \geq 2$ and these two graphs belong to distinct isomorphism classes within the same reconstruction class R_i. Finally, the number of small edges is reconstructible, i.e., $e(G_{i,j}) = e(G_{i,1})$. Therefore,

$$u^\ell(G_{i,j}) = w^{i,j}(G_{i,j}) = s(G_{i,j}, G_{i,j}) - s(G_{i,j}, G_{i,1}) = 1 - 0 = 1.$$

Now consider the vectors $u^k = w^{i',j'}$ and $u^\ell = w^{i,j}$, where $1 \leq k < \ell$. We prove that $u^k(G_{i,j}) = 0$. Since $k < \ell$, according to the ordering of U, we have $e(G_{i',j'}) \leq e(G_{i,j})$. Since $G_{i',j'}$ and $G_{i',1}$ are reconstructions of each other, we have $e(G_{i',j'}) = e(G_{i',1})$.

Now, if $e(G_{i',j'}) < e(G_{i,j})$, then

$$u^k(G_{i,j}) = w^{i',j'}(G_{i,j}) = s(G_{i,j}, G_{i',j'}) - s(G_{i,j}, G_{i',1}) = 0 - 0 = 0.$$

On the other hand, if $e(G_{i',j'}) = e(G_{i,j})$, then again $s(G_{i,j}, G_{i',j'}) = 0$ (since $G_{i,j}$ and $G_{i',j'}$ are non-isomorphic but have the same number of edges) and $s(G_{i,j}, G_{i',1}) = 0$ (because $j > 1$, so $G_{i,j}$ and $G_{i',1}$ are non-isomorphic but have the same number of edges).

Now the lemma follows from $\alpha(\mathcal{C}_n) := |\mathcal{C}_n/\!\cong| - |\mathcal{C}_n/\!\sim| = \sum_{i=1}^{s}(r_i - 1) = |U|$. □

Theorem 3.2. *Let $\mathcal{C}_n$ be a finite, recognisable class of n-vertex graphs satisfying Kocay's Lemma. Let $\mathfrak{F}$ be a family of sequences of graphs on at most $n-1$ vertices. If $M_{\mathfrak{F},\,\mathcal{C}_n/\cong}$ is the corresponding matrix of covering numbers associated with $\mathfrak{F}$ and $\mathcal{C}_n$, then $|\mathcal{C}_n/\!\sim| \geq \mathsf{rank}_{\mathbb{R}}(M_{\mathfrak{F},\,\mathcal{C}_n/\cong})$.*

Proof. Applying the Rank-Nullity Theorem, we have

$$\mathsf{dim}(W) + \mathsf{rank}_{\mathbb{R}}(M_{\mathfrak{F},\,\mathcal{C}_n/\cong}) = |\mathcal{C}_n/\!\cong|\,.$$

It follows from Lemma 3.1 that

$$\alpha(\mathcal{C}_n) + \mathsf{rank}_{\mathbb{R}}(M_{\mathfrak{F},\,\mathcal{C}_n/\cong}) \leq \mathsf{dim}(W) + \mathsf{rank}_{\mathbb{R}}(M_{\mathfrak{F},\,\mathcal{C}_n/\cong}) = |\mathcal{C}_n/\!\cong|\,.$$

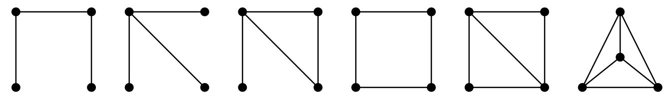
Now recalling the definition of $\alpha(\mathcal{C}_n)$, we have

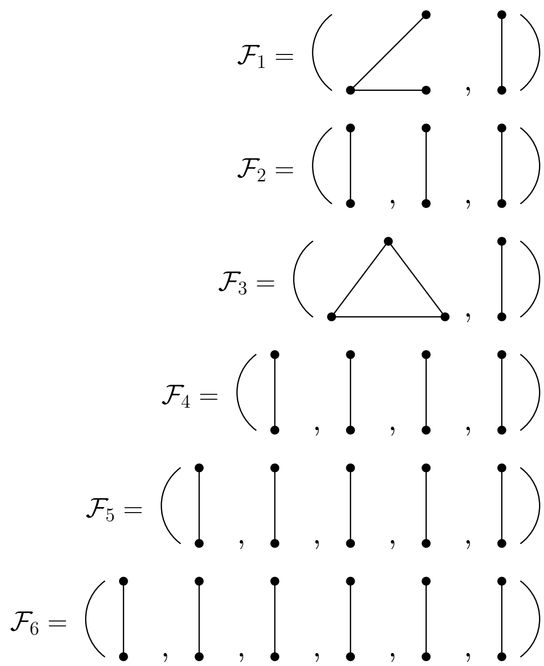
$$|\mathcal{C}_n/\!\cong| - |\mathcal{C}_n/\!\sim| + \mathsf{rank}_{\mathbb{R}}(M_{\mathfrak{F},\,\mathcal{C}_n/\cong}) \leq |\mathcal{C}_n/\!\cong|\,,$$

which implies that $|\mathcal{C}_n/\!\sim| \geq \mathsf{rank}_{\mathbb{R}}(M)_{\mathfrak{F},\,\mathcal{C}_n/\cong}$. □

Corollary 3.3. *Under the hypotheses of Theorem 3.2, if $\mathsf{rank}_{\mathbb{R}}(M_{\mathfrak{F},\mathcal{C}_n}) = |\mathcal{C}_n/\!\cong|$ then every graph in $\mathcal{C}_n$ is reconstructible.*

Figure 1 illustrates an application of Corollary 3.3 to the class of connected graphs on four vertices.





	G_1	G_2	G_3	G_4	G_5	G_6
$\mathcal{F}_1$	2	3	0	0	0	0
$\mathcal{F}_2$	6	6	0	0	0	0
$\mathcal{F}_3$	0	0	1	0	0	0
$\mathcal{F}_4$	36	36	24	24	0	0
$\mathcal{F}_5$	150	150	240	240	120	0
$\mathcal{F}_6$	540	540	1536	1536	1800	720

Figure 1: A full-rank matrix M of covering numbers $c(\mathcal{F}_i, G_j)$ providing a proof through Corollary 3.3 that all connected graphs on four vertices are reconstructible.

3.2 The existence of matrices with optimal rank

Theorem 3.4. *Let $\mathcal{C}_n$ be a recognizable class of n-vertex graphs satisfying Kocay's lemma. Then there exists a family of sequences $\mathfrak{F}$ graphs with the corresponding matrix $M_{\mathfrak{F},\mathcal{C}_n/\cong}$ of covering numbers such that $\mathsf{rank}_{\mathbb{R}}(M_{\mathfrak{F},\mathcal{C}_n/\cong}) = |\mathcal{C}_n/\!\sim\!|$.*

Proof. Let $\mathfrak{F}$ be the family of all *inequivalent* sequences of length at most n of $(n-1)$-vertex graphs. Here we consider two sequences $\mathcal{F}_i$ and $\mathcal{F}_j$ to be inequivalent if for each bijection f from $\mathcal{F}_i$ to $\mathcal{F}_j$, there is at least one graph F in $\mathcal{F}_i$ for which $f(F)$ is not isomorphic to F. Since the covering numbers for sequences of length 1 in $\mathfrak{F}$ are all 0, we assume that $\mathfrak{F}$ contains only sequences of length at least 2. Let $M_{\mathfrak{F},\mathcal{C}_n/\cong}$ be the corresponding matrix of covering numbers. We show below that this choice for the family of sequences and its corresponding matrix of covering numbers satisfy the desired property.

For a sequence $\mathcal{F}$ and a graph G, let $c^*(\mathcal{F}, G)$ denote the number of tuples $(G_1, G_2, \ldots, G_m)$ of subgraphs of G with distinct vertex sets such that $G_i \cong F_i$, $1 \le i \le m$, and $\bigcup G_i = G$. We call such covers *non-overlapping*. Correspondingly, we have the matrix $M^*_{\mathfrak{F},\mathcal{C}_n/\cong}$ of non-overlapping covering numbers.

Now let $\mathcal{F} := (F_1, F_2, \ldots, F_\ell)$ be a sequence in $\mathfrak{F}$. We have the following recurrence for $c(\mathcal{F}, G)$:

$$c(\mathcal{F},G)=\sum_{k=2}^{\ell}\sum_{P\in\mathcal{P}_\ell^k}\sum_{\mathcal{H}:=(H_1,H_2,\ldots,H_k)}\gamma(\mathcal{H})c^*(\mathcal{H},G)\prod_{i=1}^{k}c(\mathcal{F}|_{P^{-1}(i)},H_i),$$

where $\mathcal{P}_\ell^k$ denotes the set of all onto functions from $\{1,2,\ldots,\ell\}$ to $\{1,2,\ldots,k\}$, and $\mathcal{F}|_{P^{-1}(i)}$ is the subsequence of $\mathcal{F}$ consisting of $F_j; j\in P^{-1}(i)$, and the innermost sum is over all inequivalent sequences $\mathcal{H}$ of length k of graphs on $(n-1)$ vertices. This may be explained as follows. Each cover $(G_1,G_2,\ldots,G_\ell)$ of G by $\mathcal{F}$ naturally corresponds to a partition of $\{1,2,\ldots,\ell\}$ in k blocks for some $k\in[2..\ell]$, so that i,j are in the same partition if and only if graphs G_i and G_j have the same vertex set. We denote partitions of $\{1,2,\ldots,\ell\}$ in k blocks by onto maps P from $\{1,2,\ldots,\ell\}$ to $\{1,2,\ldots,k\}$ so that the inverse image $P^{-1}(i)$ denotes the i-th block. For the i-th block $P^{-1}(i)$ of an onto map P, the union of graphs $G_j; j\in P^{-1}(i)$ is a graph H_i on $n-1$ vertices. We denote the subsequence of $\mathcal{F}$ with indices $j\in P^{-1}(i)$ by $\mathcal{F}|_{P^{-1}(i)}$. Now the cover of G by the sequence $\mathcal{H}:=(H_1,H_2,\ldots,H_k)$ is non-overlapping, and each H_i may be covered by $F_j; j\in P^{-1}(i)$ in $c(\mathcal{F}|_{P^{-1}(i)},H_i)$ ways. We do not need to consider the trivial partition of $\{1,2,\ldots,\ell\}$ into a single block, because there is no cover $(G_1,G_2,\ldots,G_\ell)$ of G by $\mathcal{F}$ such that all G_i have the same vertex set. In other words, the above formula computes $c(\mathcal{F},G)$ by partitioning the coverings according to k, P, and $\mathcal{H}$, and then counting the number of coverings in each block of the partition. Since in the formula we use onto functions instead of partitions, the same block of coverings under this partition may be counted more than once, and therefore there is factor $\gamma(\mathcal{H})$ in the formula. If sequence $\mathcal{H}$ contains k_1 copies of a graph Γ_1, k_2 copies of a graph Γ_2, and so on, where Γ_i are mutually non-isomorphic graphs, then $\gamma(\mathcal{H})=(\prod_i k_i!)^{-1}$.

Now we rearrange the terms and write

$$c^*(\mathcal{F},G)=c(\mathcal{F},G)-\sum_{k=2}^{\ell-1}\sum_{P\in\mathcal{P}_\ell^k}\sum_{\mathcal{H}:=(H_1,H_2,\ldots,H_k)}\gamma(\mathcal{H})c^*(\mathcal{H},G)\prod_{i=1}^{k}c(\mathcal{F}|_{P^{-1}(i)},H_i).$$

Thus we have expressed the non-overlapping covering numbers for a sequence of length ℓ of graphs in terms of the non-overlapping covering numbers for sequences of length at most $\ell-1$. In the above equation, $c(\mathcal{F}|_{P^{-1}(i)},H_i)$ are constants independent of G. Also, if $\ell=2$, we have $c^*(\mathcal{F},G)=c(\mathcal{F},G)$. Therefore, by repeatedly applying the above equation to terms containing non-overlapping covering numbers, we eventually obtain

$$c^*(\mathcal{F},G)=\sum_{\mathcal{F}'}\beta_{\mathcal{F}}(\mathcal{F}')c(\mathcal{F}',G).$$

We have written the coefficients as $\beta_{\mathcal{F}}(\mathcal{F}')$ to emphasize that they arise from factors $c(\mathcal{F}|_{P^{-1}(i)},H_i)$ and $\gamma(\mathcal{H})$ that do not depend on G. That is, the linear dependence of the non-overlapping covering numbers on the covering numbers is the same for all graphs (but of course depends on $\mathcal{F}$). Therefore, we can write

$$c^*(\mathcal{F},-)=\sum_{\mathcal{F}'}\beta_{\mathcal{F}}(\mathcal{F}')c(\mathcal{F}',-).$$

In this manner we have shown that the rows of $M^*_{\mathfrak{F},\,\mathcal{C}_n/\cong}$ are in the span of the rows of $M_{\mathfrak{F},\,\mathcal{C}_n/\cong}$. Therefore, we have

$$\mathsf{rank}_{\mathbb{R}}(M^*_{\mathfrak{F},\,\mathcal{C}_n/\cong})\le\mathsf{rank}_{\mathbb{R}}(M_{\mathfrak{F},\,\mathcal{C}_n/\cong}).$$

To show that the rank of $M^*_{\mathfrak{F},\mathcal{C}_n/\cong}$ is $|\mathcal{C}_n/\sim|$, we construct a square submatrix K of $M^*_{\mathfrak{F},\mathcal{C}_n/\cong}$ as follows. Let $\{R_i, i = 1,2,\dots\} \coloneqq \mathcal{C}_n/\sim$. First, for each reconstruction class $R_i, i = 1,2,\dots,$ we choose one reconstruction G_i arbitrarily from $R_i/\cong$. For each $i = 1,2,\dots,$ we keep the row indexed by the sequence (say $\mathcal{F}_i$) that is equivalent to the sequence $(G_i - v, v \in V(G_i))$, where the vertices of G_i may be ordered arbitrarily, and we keep the column indexed by G_i. We delete all other rows and columns of $M^*_{\mathfrak{F},\mathcal{C}_n/\cong}$. We show that K has full rank, which will imply that $\mathsf{rank}_{\mathbb{R}}(M^*_{\mathfrak{F},\mathcal{C}_n/\cong}) \geq \mathsf{rank}_{\mathbb{R}}(K) = |\mathcal{C}_n/\sim|$.

We define a partial order $\leq$ on $\mathcal{C}_n/\sim$ so that $R_i \leq R_j$ if there exists a bijection f from $V(G_i)$ to $V(G_j)$ such that for each v in $V(G_i)$, the graph $G_i - v$ is isomorphic to a subgraph of $G_j - f(v)$.

First we verify that the above relation $\leq$ is a partial order on $\mathcal{C}_n/\sim$. The reflexivity and the transitivity are straightforward to verify. We now verify antisymmetry. Let f be a bijection as in the above paragraph. Therefore, for each $v \in V(G_i)$, we have $e(G_i - v) \leq e(G_j - f(v))$. Let g be a similar bijection from $V(G_j)$ to $V(G_i)$. Therefore, the bijective composition $g \circ f$ from $V(G_i)$ to $V(G_i)$ is such that for all v in $V(G_i)$, we have $G_i - v$ is isomorphic to a subgraph of $G_i - (g \circ f)(v)$, implying that $e(G_i - v) \leq e(G_j - f(v)) \leq e(G_i - (g \circ f)(v))$. Now observe that $\sum_v e(G_i - v) = \sum_v e(G_i - (g \circ f)(v))$, since $g \circ f$ is a bijection from $V(G_i)$ onto itself. Therefore, we must have $e(G_i - v) = e(G_j - f(v))$ for all $v \in V(G_i)$, implying that $G_i - v$ and $G_j - f(v)$ are isomorphic for all $v \in V(G_i)$. In other words, $R_i = R_j$.

We sort the rows and the columns of K so that if $R_i < R_j$, then G_j is to the right of G_i, and the row corresponding to the sequence $\mathcal{F}_i$ is above the row corresponding to the family $\mathcal{F}_j$.

Now if $c^*(\mathcal{F}_i, G_j) > 0$ then $R_i < R_j$, therefore, the matrix K is upper-triangular. Also, $c^*(\mathcal{F}_i, G_i) > 0$ for all G_i. Therefore, K has full rank; in fact $\mathsf{rank}(K)$ is equal $|\mathcal{C}_n/\sim|$. Since the class $\mathcal{C}_n$ is recognizable and satisfies Kocay's lemma, Theorem 3.2 is applicable. Therefore,

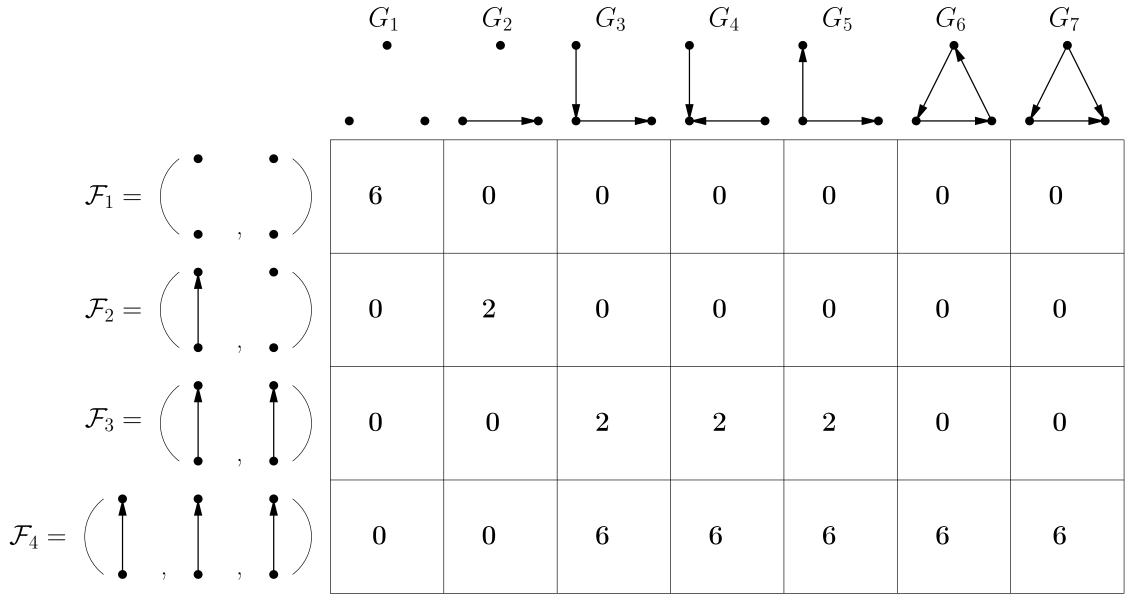
$$|\mathcal{C}_n/\sim| = \mathsf{rank}_{\mathbb{R}}(K) \leq \mathsf{rank}_{\mathbb{R}}(M^*_{\mathfrak{F},\mathcal{C}_n/\cong}) \leq \mathsf{rank}_{\mathbb{R}}(M_{\mathfrak{F},\mathcal{C}_n/\cong}) \leq |\mathcal{C}_n/\sim|\,,$$

which implies the claim for our choice of $\mathfrak{F}$, and the corresponding matrix $M_{\mathfrak{F},\mathcal{C}_n/\cong}$. □

Example 3.5. We show another small but non-trivial example in directed graphs, which are in general not reconstructible. Figure 2 illustrates a matrix of covering numbers for directed graphs on 3 vertices, with no multi-arcs or loops. Observe that there are 7 distinct graphs in 4 reconstruction classes: G_1 and G_2 are reconstructible; G_3, G_4, G_5 belong to the same reconstruction class; G_6, G_7 belong to the same reconstruction class. The figure shows 4 rows of the matrix corresponding to 4 graph sequences. The rank of the matrix is 4, which is also the number of reconstruction classes. It is possible to verify that the rank cannot be improved by adding more sequences of graphs.

4 Discussion

In this paper we have described an algebraic formulation of the graph reconstruction conjecture. Our results show that if this conjecture is true then, at least in principle, it may be proven using equations obtained from Kocay's lemma, and we believe that further investigation of this approach may be fruitful. For example, it will be interesting to prove that trees are reconstructible using the approach of this paper. On the other hand, existence of non-reconstructible directed graphs (particularly tournaments) and hypergraphs may also be proved by an algebraic approach based on Kocay's lemma.



	G_1	G_2	G_3	G_4	G_5	G_6	G_7
$\mathcal{F}_1 =$	6	0	0	0	0	0	0
$\mathcal{F}_2 =$	0	2	0	0	0	0	0
$\mathcal{F}_3 =$	0	0	2	2	2	0	0
$\mathcal{F}_4 =$	0	0	6	6	6	6	6

Figure 2: A matrix of covering numbers for directed graphs on 3 vertices. There are 4 reconstruction classes and the rank of the above matrix is also 4.

The approach may also be useful to study questions about *legitimate decks*: a collection of n-vertex graphs, each of which has $n-1$ vertices, is called a *legitimate deck* if it is the collection of vertex-deleted subgraphs of an n-vertex graph. Harary [4] asked for a characterisation of legitimate decks.

Now suppose that $\mathcal{F}$ is a sequence of n graphs, each on $(n-1)$ vertices, that is not a legitimate deck. In the proof of Theorem 3.4, we observe that a sequence equivalent to $\mathcal{F}$ does not appear in the recurrence relations for non-overlapping covering numbers of other sequences. Therefore, deleting the row corresponding to $\mathcal{F}$ (or corresponding to a sequence equivalent to it) from M does not change the rank of M. Can we extend such arguments further to completely characterise legitimate decks?

Acknowledgements

We would like to thank Hiep Han for helpful discussions. The first author is also grateful to Yoshiharu Kohayakawa for hosting him at Universidade de São Paulo.

References

[1] M. Bilinski, Y.S. Kwon, and X. Yu. On the reconstruction of planar graphs. *Journal of Combinatorial Theory, Series B*, 97(5):745–756, 2007.

[2] Béla Bollobás. Almost every graph has reconstruction number three. *J. Graph Theory*, 14(1): 1–4, 1990. ISSN 0364-9024.

[3] J. A. Bondy. A graph reconstructor's manual. In *Surveys in combinatorics, 1991 (Guildford, 1991)*, volume 166 of *London Math. Soc. Lecture Note Ser.*, pages 221–252. Cambridge Univ. Press, Cambridge, 1991.

[4] Frank Harary. The four color conjecture and other graphical diseases. In *Proof Techniques in Graph Theory (Proc. Second Ann Arbor Graph Theory Conf., Ann Arbor, Mich., 1968)*, pages 1–9. Academic Press, New York, 1969.

[5] P.J. Kelly. *On isometric transformations.* PhD thesis, University of Wisconsin–Madison, 1942.

[6] P.J. Kelly. A congruence theorem for trees. *Pacific Journal of Mathematics*, 7(1):961–968, 1957.

[7] W. L. Kocay. An extension of Kelly's lemma to spanning subgraphs. In *Proceedings of the Tenth Manitoba Conference on Numerical Mathematics and Computing, Vol. II (Winnipeg, Man., 1980)*, volume 31, pages 109–120, 1981.

[8] B.D. McKay. Small graphs are reconstructible. *Australasian Journal of Combinatorics*, 15: 123–126, 1997.

[9] Paul K. Stockmeyer. The falsity of the reconstruction conjecture for tournaments. *J. Graph Theory*, 1(1):19–25, 1977. ISSN 0364-9024.

[10] Paul K. Stockmeyer. A census of nonreconstructible digraphs. I. Six related families. *J. Combin. Theory Ser. B*, 31(2):232–239, 1981. ISSN 0095-8956.

[11] S. M. Ulam. *A collection of mathematical problems.* Interscience Tracts in Pure and Applied Mathematics, no. 8. Interscience Publishers, New York-London, 1960.

[12] D.B. West. *Introduction to graph theory.* Prentice Hall, 2001.